

DERWENT- 1998-403328
ACC-NO:

DERWENT- 199835
WEEK:

COPYRIGHT 2004 DERWENT INFORMATION LTD

TITLE: Communication method for computer network - involves transmitting data using remote memory write-in function based on position information notified by receiving side computer in packet assigned in transmitting and receiving buffer

PATENT-ASSIGNEE: HITACHI LTD[HITA]

PRIORITY-DATA: 1996JP-0331606 (November 27, 1996)

PATENT-FAMILY:

PUB-NO	PUB-DATE	LANGUAGE	PAGES	MAIN-IPC
JP 10161955 A	June 19, 1998	N/A	012	G06F 013/00

APPLICATION-DATA:

PUB-NO	APPL-DESCRIPTOR	APPL-NO	APPL-DATE
JP 10161955A	N/A	1996JP-0331606	November 27, 1996

INT-CL (IPC): G06F013/00, G06F015/163

ABSTRACTED-PUB-NO: JP 10161955A

BASIC-ABSTRACT:

The method involves initializing communication between multiple computers (101A-101N) based on user's program. A predetermined number of packets are assigned to transmitting and receiving buffer of the remote memory transmitting and receiving area (110) of each computer. During initialization, a computer at the receiving side detects data arrival from a transmitting computer.

After detection, the receiving computer assigns a packet in transmitting and receiving buffer for reception of data to be communicated and then notifies position information of the assigned packet using remote memory write-in function. Then, the transmitting side computer transmits data using remote memory data write-in function based on position information notified by the data receiving computer.

ADVANTAGE - Performs data forwarding at high speed.

CHOSEN- Dwg. 1/11
DRAWING:

TITLE- COMMUNICATE METHOD COMPUTER NETWORK TRANSMIT DATA
TERMS: REMOTE MEMORY WRITING FUNCTION BASED POSITION
INFORMATION NOTIFICATION RECEIVE SIDE COMPUTER PACKET
ASSIGN TRANSMIT RECEIVE BUFFER

DERWENT-CLASS: T01

EPI-CODES: T01-H07C5;

SECONDARY-ACC-NO:

Non-CPI Secondary Accession Numbers: N1998-314314

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-161955

(43) 公開日 平成10年(1998) 6月19日

(51) Int.Cl.⁸

G 0 6 F 13/00
15/163

識別記号

3 5 3

F I

G 0 6 F 13/00
15/16

3 5 3 Q
3 1 0 F

審査請求 未請求 請求項の数4 F D (全 12 頁)

(21) 出願番号 特願平8-331606

(22) 出願日 平成8年(1996)11月27日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 吉澤 聡

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 岩月 秀樹

神奈川県横浜市戸塚区戸塚町5030番地 株

式会社日立製作所ソフトウェア開発本部内

(72) 発明者 佐川 暢俊

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74) 代理人 弁理士 笹岡 茂 (外1名)

最終頁に続く

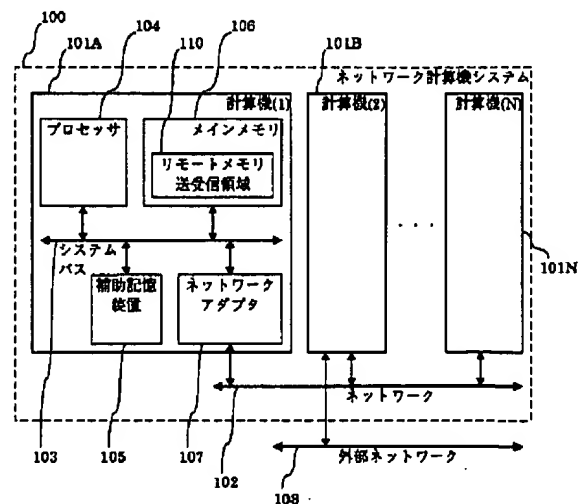
(54) 【発明の名称】 ネットワーク通信方法

(57) 【要約】

【課題】 ネットワーク計算機システムに於いて、リモートメモリ書き込み機構を用いてメッセージ転送を高速かつ安全に行うことにある。

【解決手段】 リモートメモリ書き込み機能を有するネットワーク計算機システムにおいて、各計算機101A、101B、…は、利用者プログラムからの通信初期化要求を契機として、前記リモートメモリ書き込み機能で使用する送受信バッファとして、予め定められた数のフラグメントを有する送受信バッファをリモートメモリ送受信領域110に確保し、受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対し次の通信データの受信用としての送受信バッファ内のフラグメントを割り当て、該フラグメントの位置情報を送信側計算機に対しリモートメモリ書き込み機能を用いて通知し、送信側計算機は、通知された位置情報に基づいて、リモートメモリデータ書き込み機能を用いてデータを送信する。

図 1



1

【特許請求の範囲】

【請求項1】 複数の計算機がネットワーク結合され、任意の計算機間でデータ通信を行う際には、データの送信側計算機が、受信側計算機の主記憶装置上の特定位置を指定して通信データを書き込むリモートメモリ書き込み機能を有するネットワーク計算機システムにおけるネットワーク通信方法であって、

各計算機は、利用者プログラムからの通信初期化要求を契機として、前記リモートメモリ書き込み機能で使用する送受信バッファとして、予め定められた数のフラグメントを有する送受信バッファを確保し、

受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを指定して割り当て、そのフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて通知し、

送信側計算機は、該受信側計算機により通知された位置情報に基づいて、前記リモートメモリデータ書き込み機能を用いてデータを送信することを特徴とするネットワーク通信方法。

【請求項2】 請求項1記載のネットワーク通信方法において、

前記受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを一度に複数個指定して割り当て、その複数個のフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて一度に通知することを特徴とするネットワーク通信方法。

【請求項3】 請求項2記載のネットワーク通信方法において、

前記受信側計算機は、前記一度に複数個指定して割り当てる送受信バッファのフラグメントが主記憶装置上で連続領域となるように割り当てることを特徴とするネットワーク通信方法。

【請求項4】 請求項2記載のネットワーク通信方法において、

前記受信側計算機は、前記送信側計算機に対して既に指定して割り当てた受信用バッファ内のフラグメントのうち、使用済みになっていない使用中フラグメントの数を管理し、該使用中フラグメントの数が予め定められた閾値よりも小さくなった際、該使用中フラグメントの数が該閾値よりも大きい予め定められた値となるように、複数の該フラグメントを受信用に指定して割り当てることを特徴とするネットワーク通信方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は複数の計算機をネットワークによって結合した並列計算機等に代表されるネ

2

ットワーク計算機システムに於ける計算機間のデータ通信に係わり、特にメッセージパッシングの高速性とデータの安全性の確保に関する。

【0002】

【従来の技術】 本発明に関連する技術として「オペレーティングシステムの設計II, XINUによるインターネットの構築: ダグラス・カマー著, 村井純監修, 歌代和正・酒匂順子訳, 啓学出版, 1991年11月」の21頁から24頁記載の技術がある。この従来例では送信側計算機で送信データを複数のパケットに分割し、パケット単位でネットワークに送出する。受信側計算機では、パケットを受信する毎にオペレーティング・システムに対して、割り込み処理を用いてパケットの到着を通知する。通知を受けたオペレーティング・システムは受信したパケットの先頭部分に格納された情報を参照し、到着したパケットが受信データの最終パケットであるか否かを判定する。最終パケットであった場合、通信データの受信側のユーザプログラムが受信待ち状態にあれば割り込みを用いて通信データの到着をOSが該ユーザプログラムに通知し、また該ユーザプログラムが受信待ち状態に無い場合には、該ユーザプログラムから受信要求が出される迄、OSで通信データを保持しておき、該ユーザプログラムから受信要求が出された時にそのデータをOSが該ユーザプログラムに渡す処理を行う。

【0003】 計算機をネットワークによって結合した並列計算機等に代表されるネットワーク計算機システムに於いては、上記の従来技術に加えて、通信処理速度を向上するための技術として、リモートメモリ書き込み機構がある。リモートメモリ書き込み機構では、各計算機は相手計算機のオペレーティング・システムの介入無しに、即ち、割り込み処理や、パケットのバッファリング処理を行わずに、直接相手計算機内の特定メモリ領域へのデータ転送が可能である。

【0004】 リモートメモリ書き込みを行うことのできる特定メモリ領域は、リモートメモリ送受信領域と呼ばれる。リモートメモリ書き込み機構を有するネットワーク計算機システムに於いて、システム内の異なる計算機間でメッセージパッシングによってデータの送受信を行う際の従来手法を図2を用いて説明する。

【0005】 各計算機101A, 101B間のネットワーク102を通じたデータ転送は、リモートメモリ送受信領域110A, 110B間で行われ、各計算機はリモートメモリ書き込み機構を用いて書き込みを行う相手計算機の受信バッファ位置情報222(アドレス)を保持する。先ず送信関数がコールされると、送信側のユーザプログラム中の送信データ領域201からリモートメモリ送受信領域110A内の送信バッファ211にデータがコピーされる。次にそこから受信側のリモートメモリ送受信領域110B内の受信バッファ221に、受信バッファ位置情報222を指定してリモートメモリ書き込

み機構を起動することにより、データを転送する。最後に受信関数がコールされて、受信バッファ221から受信側のユーザプログラムの受信データ領域202にデータがコピーされて、メッセージの受け渡しが完了する。

【0006】

【発明が解決しようとする課題】上述の様に、リモートメモリ書き込み機構では、相手計算機のオペレーティング・システムの介入無しに高速にデータの転送を行うため、受信側のリモートメモリ送受信領域の使用状態を確認しないと、データを損失する可能性があり、連続したデータ書き込みを安全に行うことができなかった。更に、リモートメモリ送受信領域内の受信バッファの位置及び容量が固定されているため、受信バッファの容量を超える長さのメッセージを転送する事もできなかった。本発明の目的は、任意の長さのメッセージ転送を、限られた容量のリモートメモリ送受信領域を用いて、高速かつデータの損失を起こす事無く、行う事にある。

【0007】

【課題を解決するための手段】上記課題を解決するため、本発明は、複数の計算機がネットワーク結合され、任意の計算機間でデータ通信を行う際には、データの送信側計算機が、受信側計算機の主記憶装置上の特定位置を指定して通信データを書き込むリモートメモリ書き込み機能を有するネットワーク計算機システムにおけるネットワーク通信方法であり、各計算機は、利用者プログラムからの通信初期化要求を契機として、前記リモートメモリ書き込み機能で使用する送受信バッファとして、予め定められた数のフラグメントを有する送受信バッファを確保し、受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを指定して割り当て、そのフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて通知し、送信側計算機は、該受信側計算機により通知された位置情報に基づいて、前記リモートメモリデータ書き込み機能を用いてデータを送信するようにしている。

【0008】また、前記受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを一度に複数個指定して割り当て、その複数個のフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて一度に通知するようにしている。

【0009】また、前記受信側計算機は、前記一度に複数個指定して割り当てる送受信バッファのフラグメントが主記憶装置上で連続領域となるように割り当てるようにしている。

【0010】また、前記受信側計算機は、前記送信側計算機に対して既に指定して割り当てた受信用バッファ内

のフラグメントのうち、使用済みになっていない使用中フラグメントの数を管理し、該使用中フラグメントの数が予め定められた閾値よりも小さくなった際、該使用中フラグメントの数が該閾値よりも大きい予め定められた値となるように、複数の該フラグメントを受信用に指定して割り当てるようにしている。

【0011】

【発明の実施の形態】図1は本発明の一実施例を示す、ネットワーク計算機システム100のブロック図である。ネットワーク計算機システム100は複数の計算機101をネットワーク102で結合した構成を持つ。ネットワーク計算機システム100の代表例としては、分散メモリ型の並列計算機を挙げることができる。計算機101に於いて、103は各構成ブロック間の命令及びデータの転送路であるシステムバス、104はプログラムを実行するプロセッサ、105はプログラムを記述したコード等を格納する磁気ディスク装置等の補助記憶装置、106はプログラム実行中にそのコードやデータを格納するメインメモリ、107はネットワーク102との入出力を制御するネットワークアダプタである。また、108はネットワーク計算機システム100と外部の計算機を接続するための外部ネットワークであり、100内の一つ以上の計算機101が接続されている。外部ネットワーク108を通じてネットワーク計算機システム100、端末装置や大容量ディスク装置（フエイル・サーバ）等が接続される。本発明のネットワーク計算機システム100に於いては、任意の計算機間でデータ通信を行う際に、データ送信側の計算機が、受信側計算機のメインメモリ上の特定位置を指定して、通信データを書き込むことのできる、リモートメモリ書き込み機構を有する。リモートメモリ書き込み機構で使用するメインメモリ106上の領域は、110のリモートメモリ送受信領域である。尚、図1では省略しているが、計算機101では、表示装置、キーボードやマウス等の入出力装置がシステムバス103に接続されていても良い。

【0012】図3、図4は、それぞれ送信側計算機101A、受信側計算機101Bの、メインメモリ106A、106B上に確保される領域の一実施例を示す図である。図3に示した送信側計算機101Aのメインメモリ106Aの実施例に於いては、201はユーザプログラムが送信するメッセージを格納した送信データ領域である。本発明では、リモートメモリ書き込み機構を用いたデータの送受信処理を、予め定められたデータ長で行い、この送受信処理の単位を「フラグメント」と呼ぶ。即ち、送信データ領域201の送信処理はフラグメント長毎に行い、201-1、201-2、…のフラグメントに分割して行う。301は、通信相手に対して既に送信したフラグメントの個数を格納する送信済みフラグメント通番情報であり、値「0」に初期化する。送信済みフラグメント通番情報301は通信相手毎に夫々別々に

5

用意される。リモートメモリ送受信領域110Aには、通信相手に対応して311の受信フラグメント位置情報管理テーブルが設けられる。該テーブルには#1~#jのエントリが用意される。また、受信フラグメント位置情報管理テーブル311は通信相手毎に夫々別々に用意される。また、送受信フラグメント領域321Aが設けられ、複数のフラグメント321A-1、321A-2、...、321A-kに分割される。送受信フラグメント領域321Aリモートメモリ送受信領域に1つ用意される。ユーザプログラムは、これらの領域は、ユーザプログラムの通信初期化要求等を契機として送受信フラグメント領域321A内のいくつかのフラグメントを確保する。ユーザプログラムに対応する受信フラグメント位置情報管理テーブル311の各エントリは上記初期化要求時に値「-1」に初期化され、その後確保されたフラグメントのフラグメント番号(#1~#k)がテーブル311のエントリに格納される。

【0013】図4に示した受信側計算機101Bのメインメモリ106Bの実施例に於いては、202はユーザプログラムが受信するメッセージを格納する受信データ領域である。また401は、通信相手から既に受信したフラグメントの個数を格納する受信済みフラグメント通番情報、402は、通信相手に対して既に通知済みの受信済みフラグメントの個数を格納する予約済みフラグメント通番情報である。受信済みフラグメント通番情報401、予約済みフラグメント通番情報402は通信相手毎に夫々別々に用意される。401及び402は、それぞれ値「0」に初期化する。リモートメモリ送受信領域110Bには、通信相手に対して411の受信フラグメント予約情報管理テーブルが設けられる。該テーブル411は通信相手毎に夫々別々に用意される。本テーブル411のエントリ(#1~#j)のエントリ数は、受信フラグメント位置情報管理テーブル311のエントリ数に一致させる。また、送受信フラグメント領域321Bも、送信側計算機101Aの場合と同様に設けられる。ユーザプログラムは、ユーザプログラムの通信初期化要求等を契機として送受信フラグメント領域321B内のいくつかのフラグメントを確保する。ユーザプログラムに対応する受信フラグメント予約情報管理テーブル411の各エントリは上記通信初期化要求時に値「-1」に初期化され、その後確保されたフラグメントのフラグメント番号(#1~#k)がテーブル311のエントリに格納される。図5、図6は、それぞれ送信側計算機101A、受信側計算機101Bに於ける本発明の一実施例の構成を示すブロック図である。本実施例にて示す各手段は、プログラムとして記述されプロセッサ104にて実行されるが、これら手段の一部、または全ては、例えばネットワークアダプタ107上にて実施しても良い。尚、各手段の処理内容については、後でフローチャートにて詳述する。

6

【0014】図5に示した送信側計算機101Aの実施例に於いては、510は送信データ領域201から送受信フラグメント領域321内のフラグメント321A-xにデータをコピーし、ネットワークアダプタ107を起動する、データ送信制御処理手段である。該処理手段による処理については図8のフローチャートにより説明する。530は送信済みフラグメント通番情報301と受信フラグメント位置情報管理テーブル311を参照して受信側計算機101Bの送受信フラグメント321B-yの位置情報を割り出す、受信フラグメント位置情報取得制御処理手段である。該処理手段による処理については図7のフローチャートにより説明する。550Aは、送受信フラグメント領域321Aの中で、未使用のフラグメントを管理する、未使用フラグメント管理制御処理手段である。即ち、該手段は、送信データ領域201のフラグメントのデータを送受信フラグメント領域321Aにコピーするとき、該領域321A内の未使用のフラグメントを与え、該領域321A内のフラグメントのデータが受信側に転送されたとき、該データ転送済みのフラグメントの未使用登録を受け、未使用フラグメントの管理をする。

【0015】図6に示した受信側計算機101Bの実施例に於いては、610は送受信フラグメントへのデータの到着を検知し、そこから受信データ領域202にデータをコピーする、データ受信制御処理手段である。該処理手段による処理については図9のフローチャートにより説明する。630は、受信済みフラグメント通番情報401及び予約済みフラグメント通番情報402を参照して、必要に応じて受信フラグメント予約情報管理テーブル411を更新し、リモートメモリ書き込み機構を用いてテーブル411の内容を送信側計算機101Aの受信フラグメント位置情報管理テーブル311に書き込む機能、及び送信側計算機101Aからの次の送信データが書き込まれる送受信フラグメント321B-yの位置を割り出す機能を有する、受信フラグメント予約情報取得通知制御処理手段である。該処理手段による処理については図10のフローチャートにより説明する。550Bは、送信側計算機101Aの550Aと同様に、未使用フラグメント管理制御処理手段である。即ち、該手段は、送受信フラグメント領域321B内のフラグメントのデータを受信データ領域202にコピーしたとき、該領域321B内のフラグメントの未使用登録を受け、受信フラグメント予約情報管理テーブル411に該領域321B内のフラグメントを登録する際、登録された未使用フラグメントを与え、未使用フラグメントの管理をする。

【0016】図7は、データ送信制御処理手段510の一実施例を示すフローチャートである。ステップ701で先ずデータ送信制御処理手段510内にあるカウンタNを1に初期化する。ステップ710では、未使用フラ

グメント管理制御処理手段550Aから、未使用フラグメント321A-xを取得し、ステップ720で、該取得したフラグメント321A-xに送信データ領域201のN番目のフラグメントのデータ201-Nをコピーする。ステップ730では、受信フラグメント位置情報取得制御処理手段530を実行して、受信側計算機101Bの受信フラグメントの位置情報を取得する。尚、手段530の詳細については図8のフローチャートで説明する。ステップ740では、ネットワークアダプタ107Aを起動して、上記フラグメント321A-xのデータをネットワーク102に送出、受信側計算機101Bのフラグメント321B-yに書き込む。ステップ750では、未使用フラグメント管理制御処理手段550Aに、上記データ送出の終わったフラグメント321A-xを未使用として登録する。ステップ760でカウンタNの値に1を加算し、最後にステップ770で送信データ領域201内の全てのデータの送信が完了したか否かを判定する。送信が完了していれば511で、データ送信制御処理510を終了し、完了していなければステップ710に戻り、ステップ770迄を繰り返し実行する。本実施例に於いては、フラグメントを一つずつステップ740でネットワーク102に送出したが、これは複数個づつ一括して送出することとしても良い。その場合、ステップ750でのフラグメントの未使用登録処理も、フラグメント送出後に一括して行う。ネットワーク102への送出処理を一括化することにより、ネットワークアダプタ107Aを起動することによるオーバーヘッドを軽減することが可能である。

【0017】図8は、受信フラグメント位置情報取得制御処理手段530の一実施例を示すフローチャートであり、図7のステップ730にて呼び出される。ステップ810で先ず、送信済みフラグメント通番情報301を読み込み、値を受信フラグメント位置情報取得制御処理手段530内にあるカウンタRに設定する。ステップ820では、カウンタRに設定された値に1を受信フラグメント位置情報管理テーブル311のエントリ数(j)で割った時の余りi1を加算する計算をし、その計算値番目のエントリの内容を受信フラグメント位置情報管理テーブル311から読み出して、手段530内にあるカウンタDに設定する。例えば、図3の場合、送信済みフラグメント通番情報が「0」で、カウンタRの値が「0」なら上記計算値は「1」となり、受信フラグメント位置情報管理テーブル311における計算値番目のエントリの内容は「#3」となる。ステップ830で、Dの値が「-1」に一致した場合には、ステップ820を繰り返し実行する。このときRの値は変化しない。なお、受信フラグメント位置情報管理テーブル311の内容は、図10のフローチャートで説明するように、受信フラグメント予約情報管理テーブル411に設定された内容が転送、複写される。「-1」に一致しなかった場

合には、ステップ840に進み、Rに1加算した値を送信済みフラグメント通番情報301に設定し、531でカウンタDの値をステップ730に対する戻り値として、受信フラグメント位置情報取得制御処理530を終了する。

【0018】図9は、データ受信制御処理手段610の一実施例を示すフローチャートである。ステップ901で先ず上記手段610内のカウンタNを1に初期化する。ステップ910では、受信フラグメント予約情報取得制御処理手段630を実行して、送信側計算機101Aが次に書き込む自計算機101B内の受信フラグメント321B-yの位置情報を取得する。尚、手段630の詳細については図10のフローチャートで説明する。ステップ920では、ステップ910で割り出した受信フラグメントにデータが到着しているか否かを調べる。ステップ920は、データの到着を検知する迄、繰り返し実行し、データの到着を検知した場合には、ステップ930に進む。930では、到着したデータを、受信データ領域内のN番目のフラグメントに該当する位置にコピーする。ステップ940では、未使用フラグメント管理制御処理手段550Bに、上記コピーの終わったフラグメント321B-yを未使用として登録する。なお、該魅しように登録されたフラグメントに該当する受信フラグメント予約情報管理テーブル内のフラグメントの番号は-1に更新される。

【0019】ステップ950でカウンタNの値に1を加算し、最後にステップ960で受信データ領域202内の全てのデータの受信が完了したか否かを判定する。受信が完了していれば611で、データ受信制御処理610を終了し、完了していなければステップ910に戻り、ステップ960迄を繰り返し実行する。

【0020】図10は、受信フラグメント予約情報取得通知制御処理手段630の一実施例を示すフローチャートであり、図9のステップ910にて呼び出される。ステップ1010で先ず、受信済みフラグメント通番情報401を読み込み、値を上記手段630内のカウンタSに設定する。ステップ1020では、予約済みフラグメント通番情報402を読み込み、値を上記手段630内のカウンタTに設定する。次にステップ1030でTとSの値の差を計算する。(T-S)の値は、送信側計算機101Aに対して通知してあり、かつ未だデータが到着していない受信用フラグメントの数に該当する。

【0021】(T-S)の値が予め定められた閾値LWM(Low Water Mark)を下回る場合、ステップ1060以降に進み、受信フラグメントの予約、及び通知の処理を行う。閾値LWMを下回らない場合はステップ1040に進む。ステップ1060では、未使用フラグメント管理制御処理手段550Bより未使用のフラグメントを取得する。ステップ1065では、カウンタTに設定された値を受信フラグメント予約情報管理

テーブル411のエントリ数(j)で割った時の余りに1を加算する計算をし、受信フラグメント予約情報管理テーブル411の、該計算で得られた計算値番目のエントリにステップ1060で取得したフラグメントの位置情報(例えば、#3、#5など)を設定する。ステップ1070では、カウンタTの値を1加算し、再度(T-S)を計算する。(T-S)の値が、予め定められた閾値MAXよりも小さい場合には、ステップ1060に戻り、ステップ1075迄を繰り返して実行する。ここで閾値MAXの値は、送信側計算機101Aに対して通知しており、かつ未だデータが到着していない受信用フラグメントの最大許容数に相当する。また閾値MAXの値は、受信フラグメント予約情報管理テーブル411のエントリ数を超えない値である。ステップ1075で、(T-S)の値が閾値MAXよりも小さくなった場合には、ステップ1080に進み、受信フラグメント予約情報管理テーブル411を、リモートメモリ書き込み機構を用いて、送信側計算機101Aの受信フラグメント位置情報管理テーブル311に書き込む。ステップ1085では、カウンタTの値を、予約済みフラグメント通番情報402に設定し、ステップ1040に進む。ステップ1030で、(T-S)の値が閾値LWMを下回らない場合にも、ステップ1040に進む。ステップ1040では、カウンタSに設定された値を受信フラグメント予約情報管理テーブル411のエントリ数(j)で割った時の余りに1を加算する計算をし、その計算値番目のエントリの内容を受信フラグメント予約情報管理テーブル411から読み出して、上記手段630内のカウンタBに設定する。ステップ1050で、カウンタSの値に1加算して、受信済みフラグメント通番情報401に設定する。631で、カウンタBの値をステップ910に対する戻り値として、受信フラグメント予約情報取得通知制御処理630を終了する。

【0022】図11に送受信フラグメント領域321内の各フラグメントの使用状態を管理する方式の一実施例を示す。1110は送受信フラグメント管理テーブルであり、1110の各エントリは、送受信フラグメント領域321内の各フラグメント321-1、・・・、321-kに対応する。1120は未使用フラグメント管理情報であり、現在未使用のフラグメントに対応する送受信フラグメント管理テーブル1110のエントリのチェーンの先頭を指し示す。1130は使用中フラグメント管理情報であり、現在使用中のフラグメントに対応する送受信フラグメント管理テーブル1110のエントリのチェーンの先頭を指し示す。送受信フラグメント管理テーブル1110の全てのエントリは、必ず上記何れか片方のチェーンに属する。尚本実施例では、テーブル1110、1120共に、チェーンの先頭のみを指し示したが、チェーンの終端も指し示す様に実施しても良い。

【0023】

【発明の効果】本発明によれば、複数の計算機をネットワークによって結合した並列計算機等に代表されるネットワーク計算機システムに於いて、送受信バッファの割り当てを、受信側から送信処理に先立ち必要に応じて行うので、リモートメモリ書き込み機構を用いて、任意の長さのメッセージ転送を、限られた容量のリモートメモリ送受信領域で、高速かつデータの損失を起こす事無く、行うことができる。

【図面の簡単な説明】

10 【図1】本発明の一実施例のネットワーク計算機システムのブロック図である。

【図2】ネットワーク計算機システムに於ける通信制御方式の従来例を示すブロック図である。

【図3】本発明のネットワーク計算機システムに於ける、送信側計算機の一実施例を示す説明図である。

【図4】本発明のネットワーク計算機システムに於ける、受信側計算機の一実施例を示す説明図である。

20 【図5】本発明のネットワーク計算機システムに於ける、送信側計算機の一実施例の構成を示すブロック図である。

【図6】本発明のネットワーク計算機システムに於ける、受信側計算機の一実施例の構成を示すブロック図である。

【図7】本発明のデータ送信制御処理部の一実施例を示すフローチャートである。

【図8】本発明の受信フラグメント位置情報取得制御処理部の一実施例を示すフローチャートである。

【図9】本発明のデータ受信制御処理部の一実施例を示すフローチャートである。

30 【図10】本発明の受信フラグメント予約情報取得通知制御処理部の一実施例を示すフローチャートである。

【図11】本発明の送受信フラグメント管理情報の一実施例を示すブロック図である。

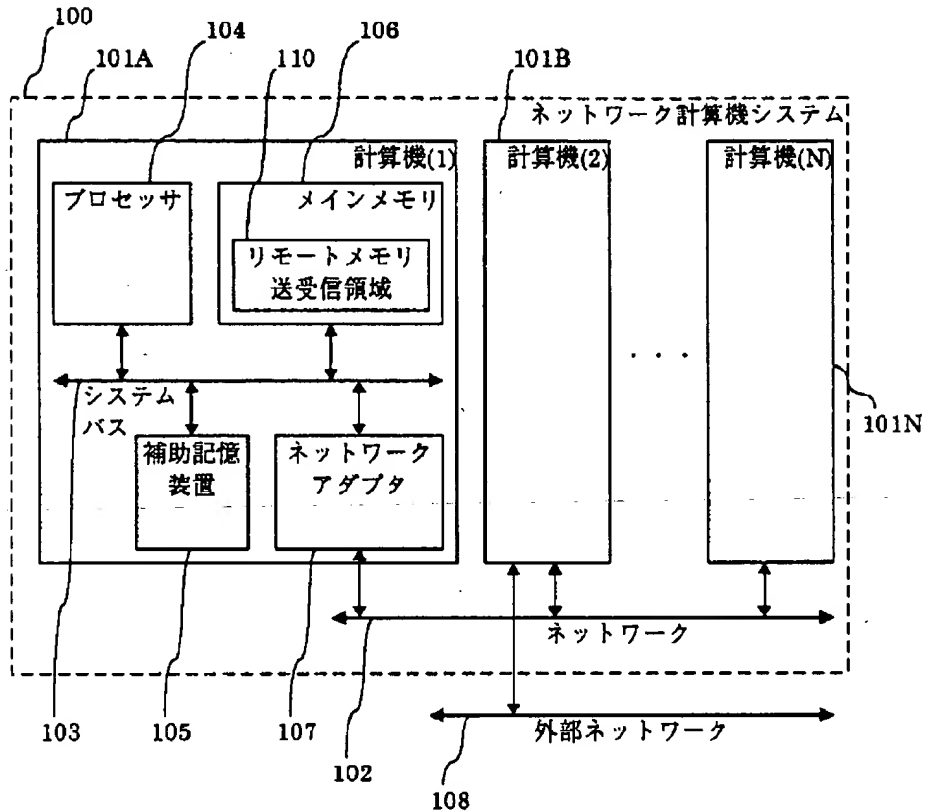
【符号の説明】

- 100 ネットワーク計算機システム
- 101 計算機
- 102 ネットワーク
- 103 システム・バス
- 104 プロセッサ
- 40 105 補助記憶装置
- 106 メインメモリ
- 107 ネットワークアダプタ
- 108 外部ネットワーク
- 110 リモートメモリ送受信領域
- 201 送信データ領域
- 202 受信データ領域
- 301 送信済みフラグメント通番情報
- 311 受信フラグメント位置情報管理テーブル
- 321 送受信フラグメント領域
- 50 401 受信済みフラグメント通番情報

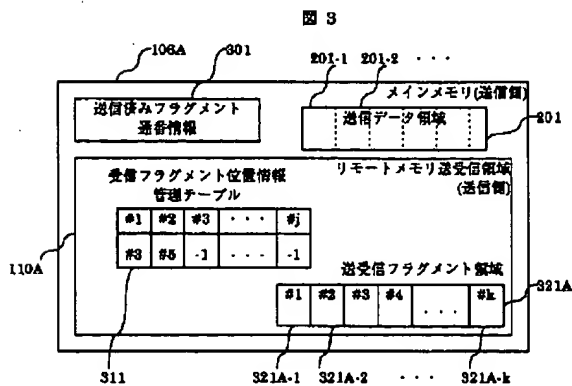
- 11
- 402 予約済みフラグメント通番情報
 411 受信フラグメント予約情報管理テーブル
 510 データ送信制御処理手段
 530 受信フラグメント位置情報取得制御処理手段
 550 未使用フラグメント管理制御処理手段
 610 データ受信制御処理手段
- 12
- 630 受信フラグメント予約情報取得通知制御処理手段
 701-1085 処理ステップ
 1110 送受信フラグメント管理テーブル
 1120 未使用フラグメント管理情報
 1130 使用中フラグメント管理情報

【図1】

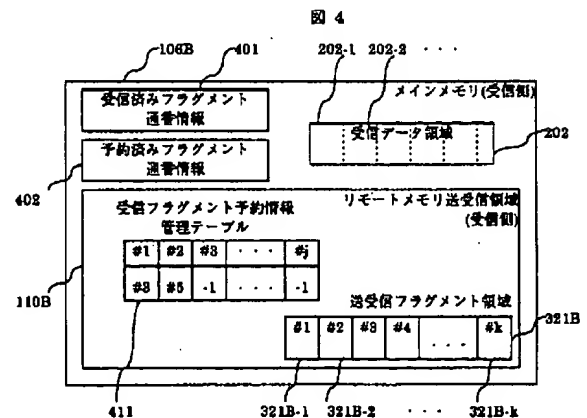
図 1



【図3】

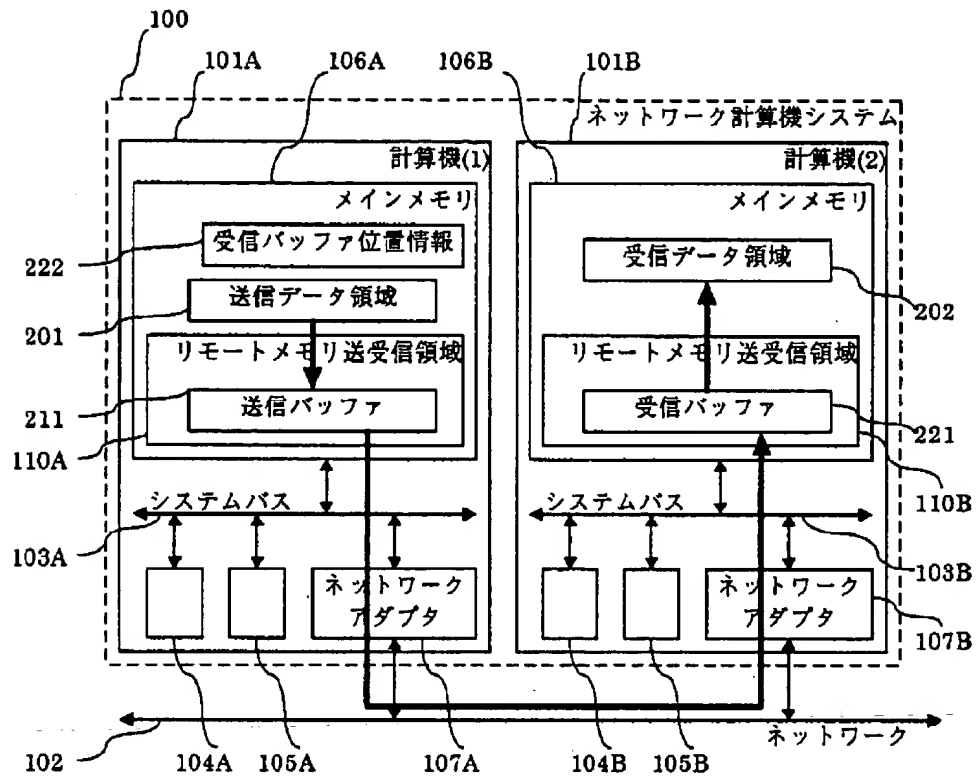


【図4】



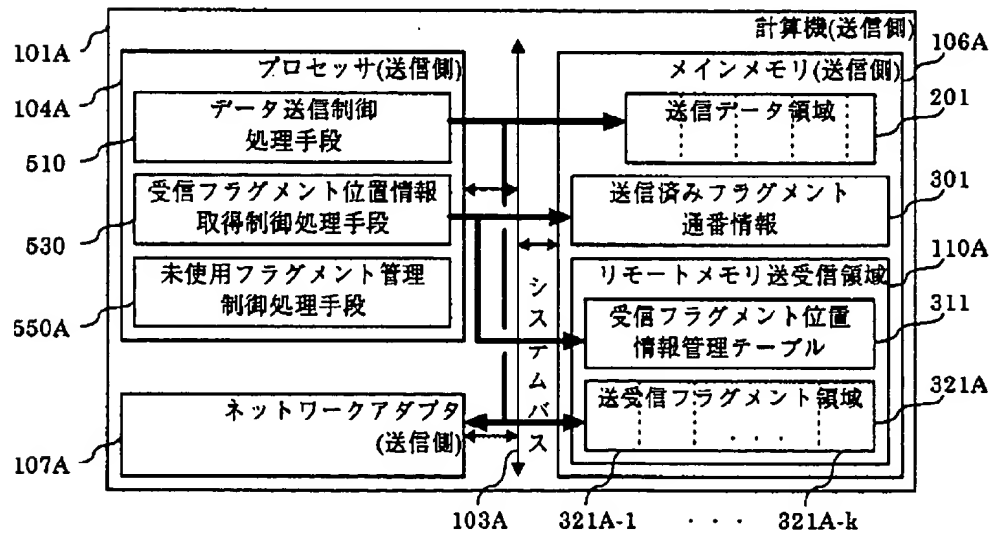
【図2】

図 2



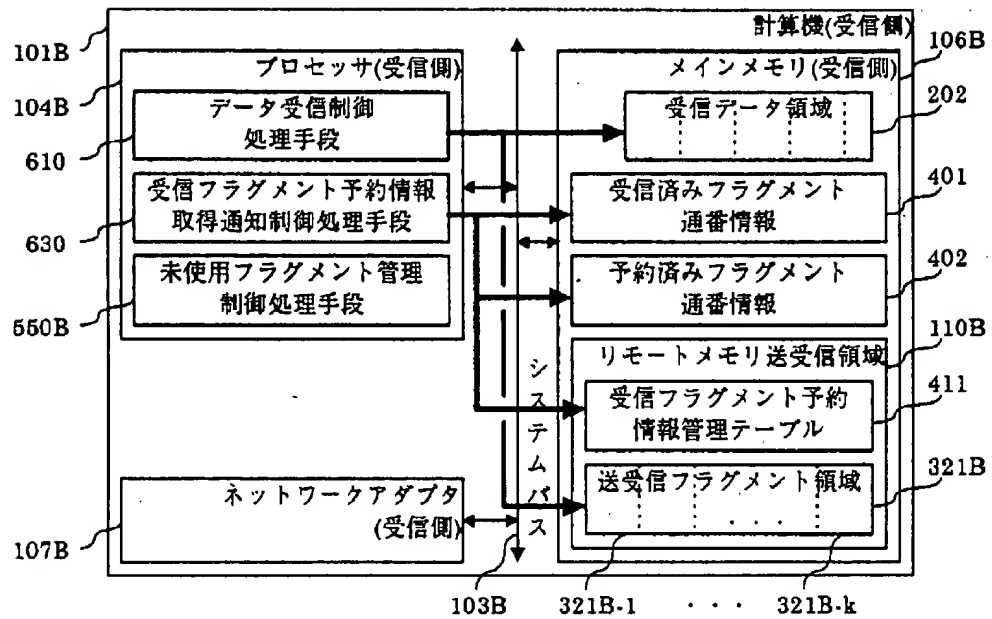
【図5】

図 5



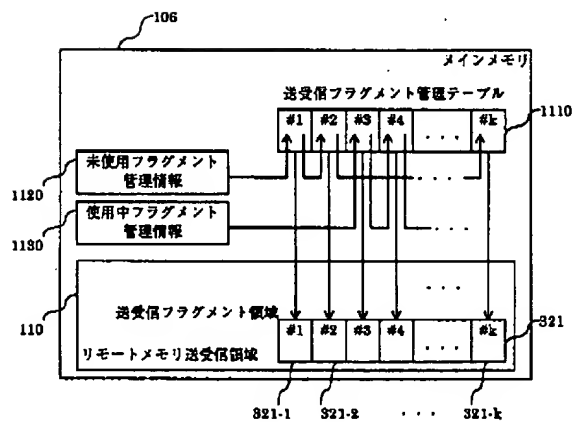
【図6】

図 6



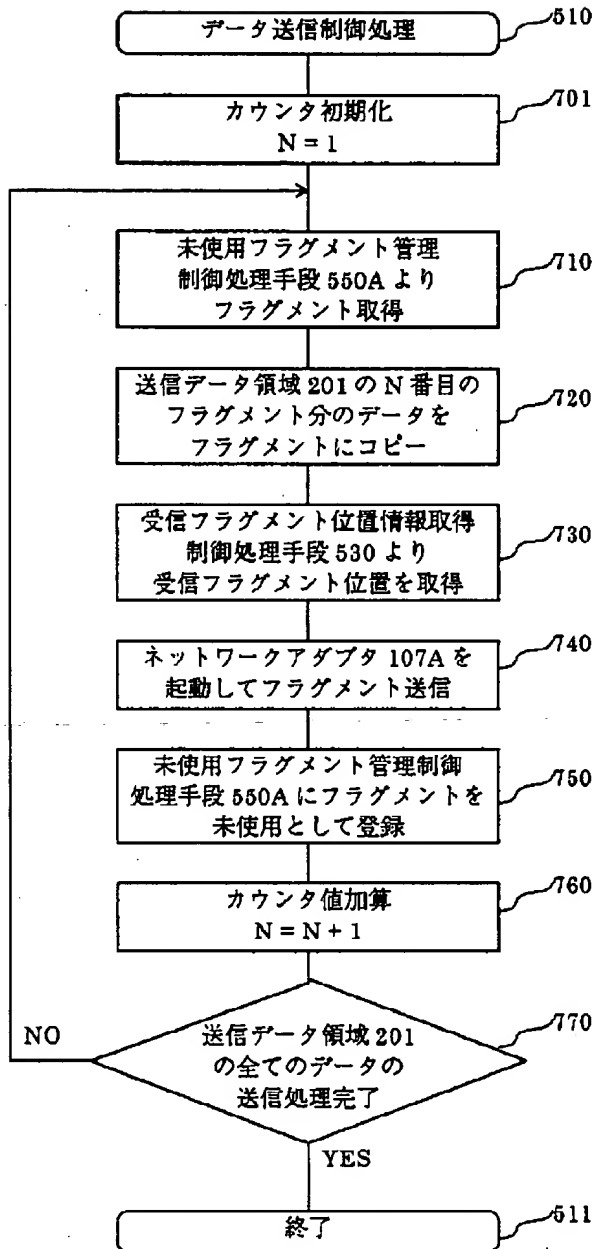
【図11】

図 11



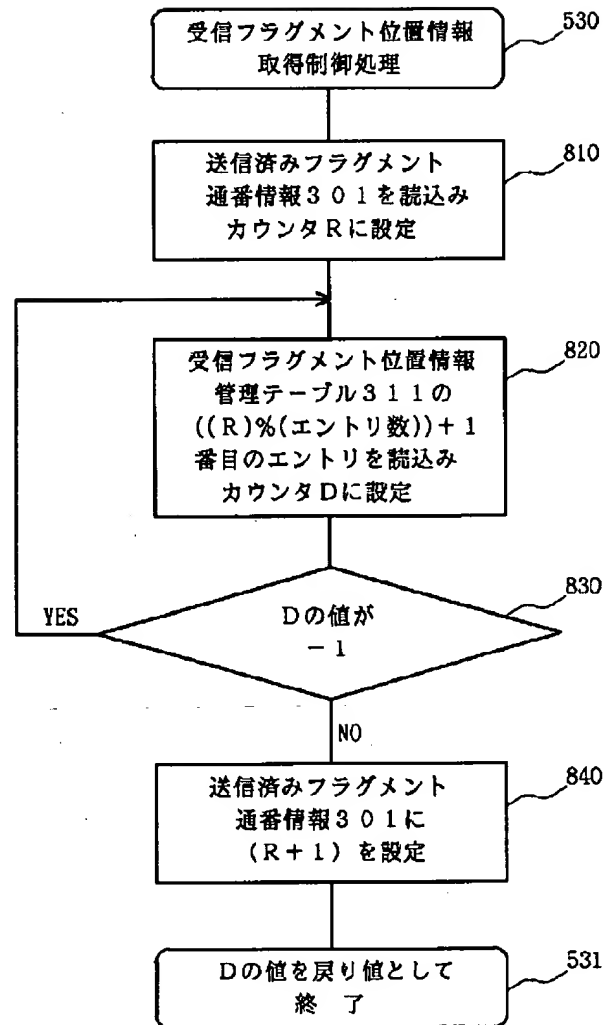
【図7】

図 7

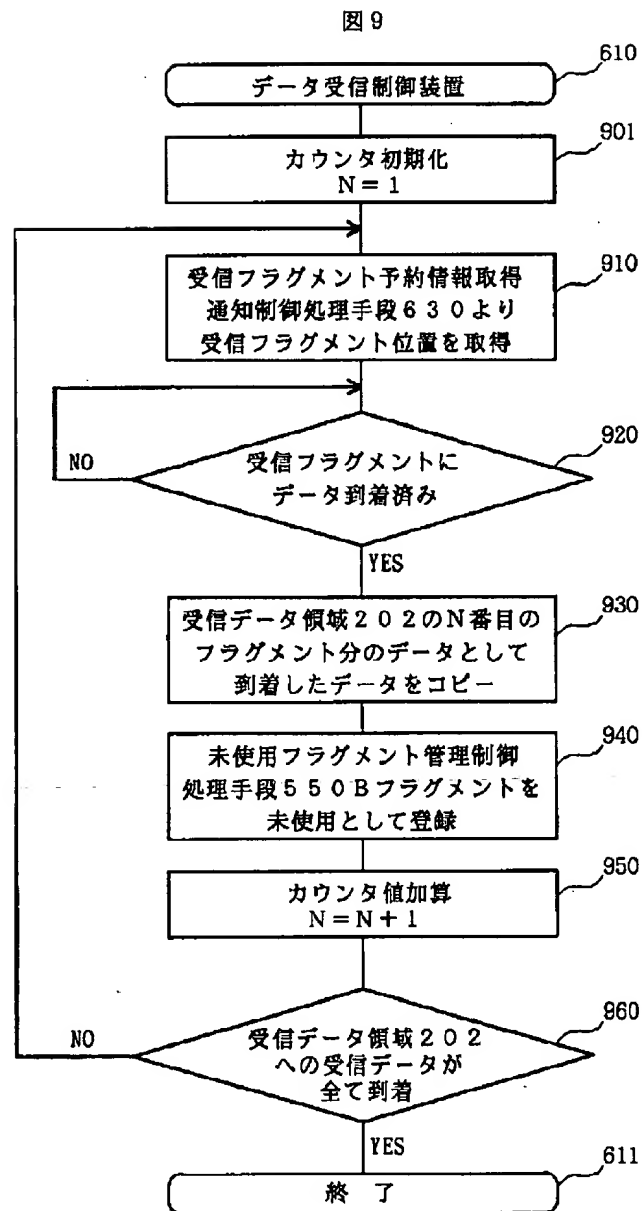


【図8】

図 8

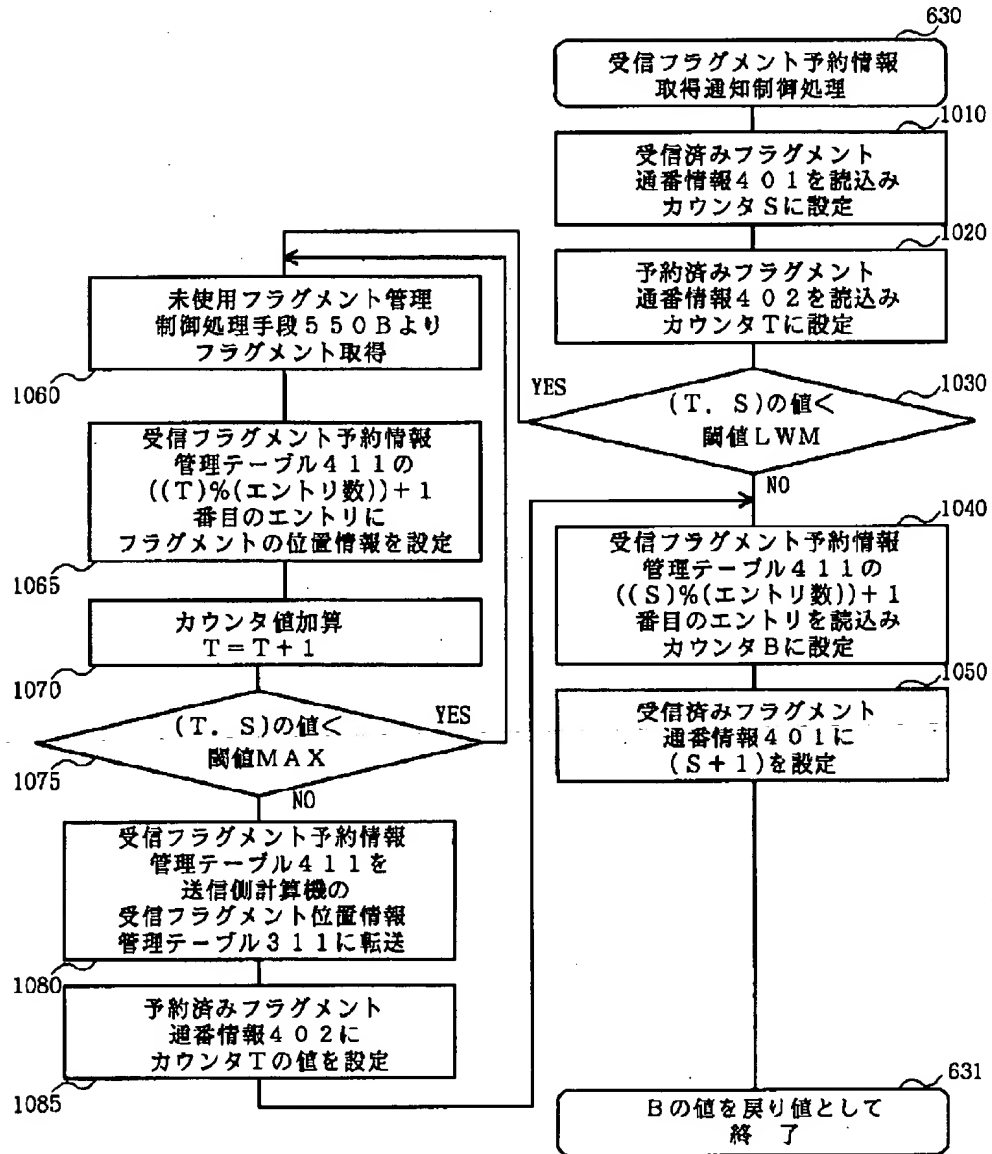


【図9】



【図10】

図10



フロントページの続き

(72)発明者 坂口 明彦
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者 今木 常之
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内